Structural Systematics in Molecular Inorganic Chemistry

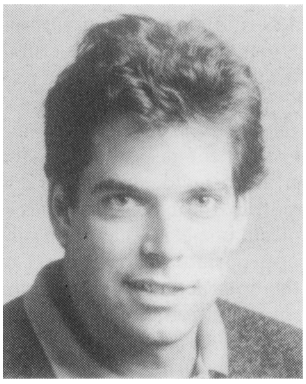
A. Guy Orpen
School of Chemistry, University of Bristol, Bristol BS8 1TS

1 Introduction

The ability crystallography gives us to visualize the world at the atomic level has revolutionized the way we look at chemistry. From the first structure determination – of NaCl – to the current time, our whole view of the gamut of chemical structure, from the simplest ionic arrays to the most complex biomolecules, has been shaped by the results of crystallography. The diversity of molecular structure is perhaps most notable in the field of transition metal chemistry where many of the traditional tenets of organic structural chemistry have had to be repeatedly re-evaluated. Because of the clarity and unequivocal answers it offers, in large areas of modern chemistry *X*-ray crystal structure analysis is often the analytical method of choice. In many laboratories, when a new organometallic or coordination complex is prepared its *X*-ray crystal structure is determined for the sake of unambiguous identification (and to expedite publication!). Often this is an end in itself and the level of detail provided by the structure analysis sufficient to provide insight into the properties of the compound, be they spectroscopic, or reactivity, or otherwise. This article, however, addresses other applications of structural data to chemistry which profit from the maturity of the techniques of crystal structure determination.

The advent of computer-controlled diffractometers and the development of increasingly powerful software and hardware packages for the analysis of diffraction data from single crystals has led to an explosive growth in available crystal structure data. Fortunately, crystallographers have made great efforts to provide these data in readily useable, computer accessible form. The largest collection of crystallographic data on molecular structure is that contained in the Cambridge Structural Database (CSD).[1] This database is devoted to compounds of 'organic carbon'. In practice about 50% of the over 100 000 crystal structures it currently (January 1993) contains fall within the domain of molecular inorganic chemistry. It therefore holds nearly *all* the crystal structures of complexes of the transition metals. This is a colossal store of structural information on the chemistry of these elements and offers opportunities for whole new areas of research in which collections of structures, rather than individuals, are the subject matter of the investigation.

Guy Orpen was born in 1955 on the island of Trinidad. He was educated in the UK and subsequently obtained his B.Sc. at the University of Cape Town where he worked with Dr. L. R. Nassimbeni. He studied for his Ph.D. at Cambridge University under the guidance of Dr. G. M. Sheldrick and Professor J. Lewis, during which time he was awarded a Walter C. Hamilton Scholarship to carry out neutron diffraction research at Brookhaven National Laboratory, USA. In 1979 he was appointed to a lectureship in Inorganic Chemistry at the University of Bristol, and in 1990 was promoted to Reader in Structural Chemistry. He has been awarded the Meldola and the Corday-Morgan Medals of the Royal Society of Chemistry. In 1992 he held a Ciba-Geigy Senior Research Fellowship. He is author and co-author of over 200 research papers in the field of structural chemistry.



The data contained within the CSD or other crystallographic databases can be of value in a number of ways. At the simplest level, literature surveys may be carried out, or compounds related to one currently under study may be retrieved for purposes of comparison. An alternative approach is to examine the geometry of a molecular or sub-molecular fragment in the variety of environments in which it has been found in crystals. This systematic type of study, as applied to problems in coordination and organometallic chemistry, is the subject of this article. At the heart of such systematic studies is enquiry into the relationship between the molecular connectivity and the shape(s) that result from it, and the properties that in turn result from the molecular shape.

2 Systematic Studies

Of the many possible applications of structural data to inorganic chemistry this article focuses on four. Examples will be taken from our own work as well as those of others in the field but the reader should be aware that this is very far from a full overview of this area of structural chemistry and is referred to more complete recent surveys.[2]

3 Typical Molecular Dimensions

One major application of the CSD and other databases has been to provide information on 'typical' dimensions for molecular parameters, such as bond lengths and angles. The function of such information is to provide a knowledge base to support efforts to build realistic structural models for molecular species in cases where direct observation of the geometry is either not practical or impossible. Thus reliable standard starting geometries are invaluable as starting points for quantum mechanical optimization of a molecular geometry, for use in modelling the interaction of a small molecule with a surface or enzyme, for use in a constrained or restrained crystallographic refinement, or for model-fitting of *e.g.* EXAFS data.

The question of what is a 'typical' geometry arises immediately. In principle, one can conceive of two approaches. First, the structure analysis of simple archetypal molecules (methane, benzene, water, $Ni(CO)_4$) provides precise dimensions for C–H, C–C bonds *etc.* in well defined environments. The second alternative is to examine the corresponding bond type or functional group in a wider range of environments, thus exploiting a database of all the relevant structural data. The advantage of the second type of study is that the bond length (for example) under study is observed under a variety of molecular and crystal environments. This allows for some estimate of the effects of changes in the environment on the molecular parameter under study, as well as those due to experimental uncertainties. Thus both the mean value of a parameter *and* its variability are important pieces of information and both need to be thought about in the discussion of suitably 'typical' dimensions for a structural unit.

In two large studies of this type we have compiled tables of bond lengths from the CSD for organic[3] and organometallic compounds, and coordination complexes of the transition elements.[4] These compilations are a summary of a large body of structural information and therefore also of a large body of chemistry. In the latter table the information is presented according to ligand and covers both metal–ligand and intraligand bond distances The chemical content is clear Thus we see how metal–phosphorus bond lengths in phosphite complexes differ from those for related phosphines, or how alkylisocyanides and the cyanide ligand differ in dimensions In addition, information on the effects on bond distances of metal oxidation state or coordination number and estimates of the variability of the bond lengths are provided, where possible One clear message of these estimates is the much wider variation of bond lengths in the transition metal study than for the organic table In part this reflects the softness of metal–ligand bonds, *i e* their sensitivity to their chemical and crystal environment In addition it reflects unresolved chemical information, in the sense of other factors not always explicitly taken into account in the tabulation (*trans* influence, coordination geometry, *etc*) Clearly these tabulations have thrown numerous areas ripe for future study, as of the type described below

4 Testing Theories

One particularly powerful application of structural data is in providing experimental tests of theories Such theories might range from qualitative, which for example might predict the extension of a bond length as a result of contraction of a bond angle, to quantitative, in which absolute values of structural parameters are obtained In either case, tests may be formulated and where possible evidence located in the CSD to allow for experimental evaluation of the theory or hypothesis While it is not possible to *prove* any scientific theory by testing it against experimental observations, it is of value to disprove those that fail to provide adequate explanations of observations [5]

Metal–ligand bonding is a particularly important field of theory for practical transition metal chemistry since ligands provide a powerful and subtle means of controlling reactivity, catalytic, and other properties of the metal Tertiary phosphine and related ligands [PA_3 (A = alkyl, aryl, OR, *etc*)] have been employed in this way and have attracted much attention from synthetic and theoretical chemists as a result Current theories of the metal–phosphine bond view it as comprising σ and π components, the latter being of debatable and certainly variable importance The nature of the π acceptor function has also been a matter of dispute, with phosphorus $3d$ or P–A σ^* orbitals, or a combination of the two, being suggested (see Figure 1) Using two different approaches we have investigated these hypotheses In the first we examined the molecular structures of pairs of transition metal phosphine complexes (L_nM-PA_3) related by redox couples [6] We looked for changes in M–P and P–A bond lengths and A–P–A bond angles, as a consequence of metal-centred oxidation The changes observed are strikingly consistent M–P distances increased, P–A distances decreased, and A–P–A angles increased These effects were entirely consistent with significant M–P π back-bonding in these complexes and with a notable P–A σ^* component in the PA_3 π acceptor function

While this kind of study is a powerful test of theories of bonding and electronic structure[7] it lacks generality, not least because of the difficulties associated with preparing crystals of pairs of complexes[1] More seriously, not all theories are testable by such a method For example, the redox-pair study said little about the effects of M–P σ bonding in these complexes We therefore set out to conduct a more general study on triphenylphosphine species ($Z-PPh_3$, Z = metal, C, O, N, *etc*)[8] in which we concentrated primarily on symmetrical distortions (*i e* preserving C_{3v} local symmetry of the ZPC_3 unit) as a function of the element Z and the Z–P bond length This study provided further substantial support for (but as always, no proof of!) a model of the bonding of triphenylphosphine in which σ effects are largely dominant, but π effects on geometry of the sort described above[6] are notable The σ effects are especially important when Z = *p*-block element (O, N, C, *etc*), while for the transition elements, σ and π effects are broadly in balance leading to PPh_3 geometries close to that of PPh_3 itself In the study of chiral complexes [$CpM(XO)Z(PPh_3)$] we obtained evidence for the asymmetry of π back-bonding between metal and phosphine being sufficient to cause observable P–C bond length and C–P–C bond angle distortions [9]

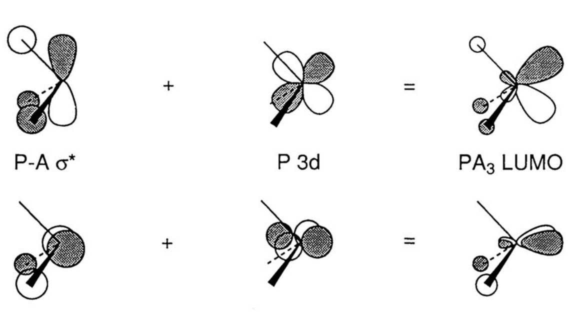


Figure 1 Hybridized π acceptor orbitals on tertiary phosphine and related ligands PA_3

5 Conformational Analysis

Flexible molecules present a particular challenge to the structural chemist An individual structure determination tells us in some sense what the average (or possibly the equilibrium) geometry of a given species is in the state in which the experiment was made (*e g* the crystal in this article) However, it is more difficult to learn much about the motion of the molecule far from its equilibrium geometry (vibrational motion close to the equilibrium geometry is, however, amenable to study by crystallographic methods – see, *e g* , reference 10) How then can we proceed when faced with molecules and ligands which we know to be flexible (or even fluxional) in solution? One approach that offers promise is to select a molecular or sub-molecular fragment and examine its shape in a (large) number of crystal and molecular environments These environments can be expected to perturb the geometry of the fragment and allow inspection of the distortions which result In the case where no bonds are broken or formed as a result of the perturbations, these distortions are variations in the *conformation* of the fragment and may allow insight into

(a) the types of conformation possible for the fragment,
(b) how these conformers might interconvert,
(c) which aspects of the molecular environment are responsible for the particular conformation adopted

These problems are the concerns of conformational analysis, and their resolution is of central importance in a range of contemporary chemistry (*e g* drug design, homogeneous catalyst design) in which the shape (or range of possible shapes) of a flexible molecule may be critical to its properties

Underlying this approach is the idea that the potential energy hypersurface (PEH) of the fragment may be read qualitatively (but *not* quantitatively)[11] from the distribution of geometries observed We assume in this kind of study that the fragments adopt geometries that are relatively low in energy and therefore map out the low-lying regions of their common PEH While for molecular fragments it is literally the molecular PEH that is being probed, for sub-molecular fragments we are not inspecting a single PEH, but rather the common features of the various PEHs of the many different molecules in the set of structures studied It is clear that studies of this sort are bound to lead to a fuller understanding of the flexibility of a fragment than are individual structure analyses In a sense the range of crystal environments to which a chosen fragment is exposed in the structures studied mirrors the variation in local fields that a molecule might experience in solution Although this analogy is in general more applicable for a molecular species than for a sub-molecular fragment, it is clearly appropriate in the context of conformational analysis where the interactions between the fragment and its environment are, by definition not so strong as to cause bond cleavage or formation

We and others have studied the conformations of both organic and inorganic systems in this way Thus when planning syntheses of extended saturated carbocyclic systems containing connected quaternary centres we wanted to know what conformations were likely to be adopted.[12] The answer turned out to be surprisingly simple. We inspected the geometries of the many tetraalkyl ammonium salts whose crystal structures are known, since they of course contain quaternary centres attached to saturated carbon chains. All NEt_4^+ ions adopted one of only two conformers which have molecular symmetry close to D_{2d} and S_4 respectively [see Figure 2 (a), (b)]. The same conformation about N–C bonds was observed for all the $\mathrm{NPr}^n_4{}^+$ and $\mathrm{NBu}^n_4{}^+$ ions. In addition *all* N–C–C–C torsion angles were close to 180°, [see Figure 2 (c)], indicating that the quaternary centre controls the conformation of three carbon chains. In contrast both *gauche* and *anti* conformations were observed in the C–C–C–C chains of the $\mathrm{NBu}^n{}_4^+$ ions. These observations are in excellent accord with calculated (molecular mechanics) energies for the observed and other unobserved conformers. As a result we were able to predict two possible high symmetry polymeric carbocyclic structures, made up of fused sixteen-membered rings of very similar energies. In the first, where the quarternary centres adopt the D_{2d} conformation, a two-dimensional molecular mat is formed, and in the other, in which the quarternary centres have the S_4 conformation, a three-dimensional cage structure related to that of the ultramarines is obtained.[12]

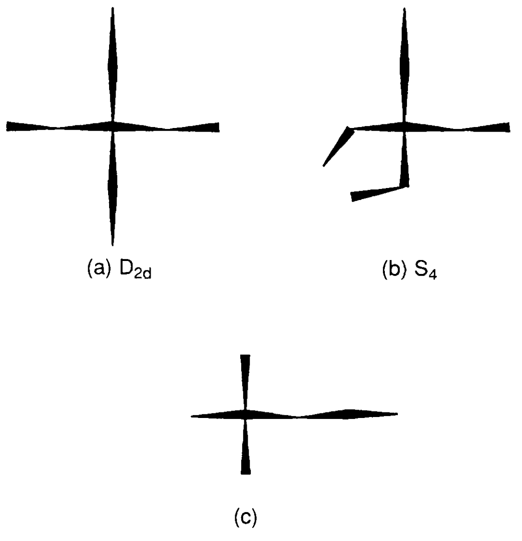


Figure 2 Preferred conformations near the quarternary centres in NEt_4^+, $\mathrm{NPr}^n_4{}^+$, and $\mathrm{NBu}^n_4{}^+$. Two distinct local symmetries are observed at the nitrogen: (a) D_{2d} and (b) S_4. All N–C–C–C conformations are *anti* (c).

The conformations of ring systems have been much studied in both organic and inorganic chemistries. One key problem is how to describe the ring conformation succinctly using the minimum number of parameters, and preferably ones that may be easily related to familiar intramolecular parameters such as torsion or dihedral angles. In our work in this area we have used the technique of principal component analysis (pca) to achieve these ends. Inspecting and understanding the torsion angle data which can be used to describe the conformations of, *e.g.*, a five-membered ring is rendered awkward because there are five such torsion angles per ring. Thus a scatterplot of such a dataset would need, on the face of it, to be five-dimensional! In practice things are not as bad as they seem, since an n-membered ring is well known to have only $n - 3$ degrees of conformational freedom (*i.e.* only two parameters or graphical axes are needed to display all the variations in conformations of a five-membered ring). However, the choice of the 'best' $n - 3$ parameters has been the subject of much discussion and research. Pca provides an objective means of reducing the dimensionality of a dataset (of any type) by extracting new axes on which to display the data which are linear combinations of the original parameters. The new axes – the principal components – are orthogonal and are chosen such that the first such component (pc1) expresses the largest fraction of the total variance of the dataset, pc2 the next largest, and so on for as many pcs as are required to describe all the variance in the data set. Figure 3 illustrates such an analysis for a dataset that is nearly (but not quite) one dimensional. A second critical element of such studies is the correct treatment of the symmetry of the parameter space, for which a variety of treatments have been considered. For torsion angle data these essentially boil down to one of two approaches.

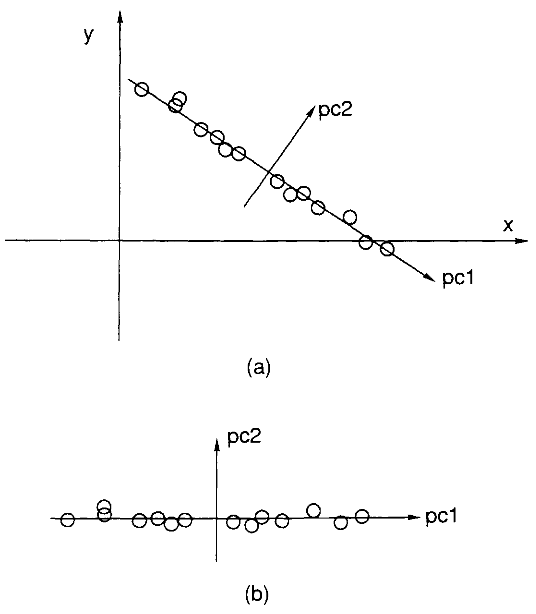


Figure 3 Schematic of principal component analysis for a near one-dimensional dataset. A plot on axes corresponding to initial parameters x and y (a) is converted to one on new axes on pc1 and pc2, which are linear combinations of x and y (b). In these plots (and those of Figures 5, 6, 7, 9, and 10) each point represents an individual structure.

In the first the initial set of torsion angle values is symmetry expanded to fill conformation space with the set of values corresponding to isometric conformations, while in the second all the conformations are symmetry transformed to lie in the asymmetric unit of conformation space.[13,14]

The graphical identification of favoured conformations is greatly aided by pca of a torsion angle dataset since by definition it produces the most helpful, *i.e.* most spread out, scatterplots of the dataset. Thus in one study of this sort we examined structures containing $\mathrm{M}_2(\mu\text{-dppm})$ [M = transition metal, μ-dppm = bridging bis(diphenylphosphino)methane] fragments or M(dppe) [dppe = chelating bis(diphenylphosphino)ethane] fragments (see Figure 4).[15] We examined the conformational preferences of the five-membered $\mathrm{MP}_2\mathrm{C}_2$ (or $\mathrm{M}_2\mathrm{P}_2\mathrm{C}$) rings and those of the attached phenyl groups. The M(dppe) fragments showed a preference for a twist (C_2) conformation of the $\mathrm{MP}_2\mathrm{C}_2$ five-membered ring (δ and λ, see Figure 5), in which the P–C–C–P torsion angle was far from zero (typically *ca.* ± 50°). In contrast, the $\mathrm{M}_2\mathrm{P}_2\mathrm{C}$ five-membered rings of the $\mathrm{M}_2(\mu\text{-dppm})$ predominantly had envelope (C_s) conformations, with the methylene carbon out of the plane of the near planar $\mathrm{M}_2\mathrm{P}_2$ unit, *i.e.* having the PMMP torsion angle close to zero, and the MPCP torsion angles *ca.* ± 45°. In these two systems the pcs generated by pca can be identified as measuring the 'twistiness' and 'envelopeness' of the conformation (*i.e.* the components of deviation from planarity along coordinates preserving C_2 or C_s symmetry of the ring system, respectively). The distribution of structures indicated that in both cases the ring conformations interconvert by a pseudo-rotation pathway, similar to those

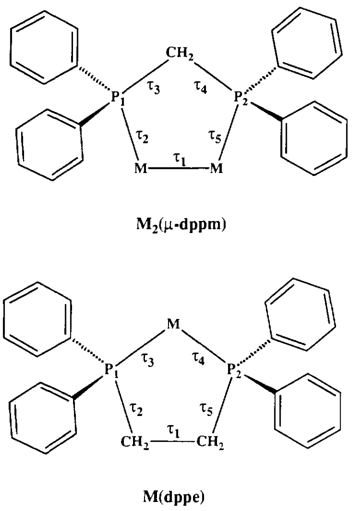


Figure 4 Torsion angles within the five-membered ring of $M_2(\mu$-dppm) and M(dppe) fragments used in pca (see Figure 5).

observed for other five-membered ring systems, in which planar intermediates (at the centre of the scatterplot in Figure 5) are avoided. The phenyl group orientations were indicative of a strong coupling of the rotations of the two phenyls on any one phosphorus in either M(dppe) or $M_2(\mu$-dppm) fragments. For the $M_2(\mu$-dppm) system transannular phenyl group interactions which lead to preferred conformations were seen. This was most pronounced for the envelope conformation of the five-membered ring, and resulted in the two pseudo-axial phenyls being near parallel. In contrast transannular interactions appeared to be insignificant for the MM(dppe) system, with little or no sign of preferred conformations for the phenyl groups.

Pca really comes into its own when dealing with more complex ring systems for which there is no analytical procedure analogous, for example, to the Cremer–Pople method.[14b,16] Thus, for example, pca of the ten intra-ring torsion angles in diethylenetriamine complexes [L_nM(dien)] gives a pca scatter-plot as shown in Figure 6.[17] The two most important pcs display 89% of the total variance in the dataset. The plot clearly shows there to be four main conformer types (δδ and λλ, λδ, δλ). Of these only δλ is observed for *mer*-M(dien) species while δδ (and its mirror

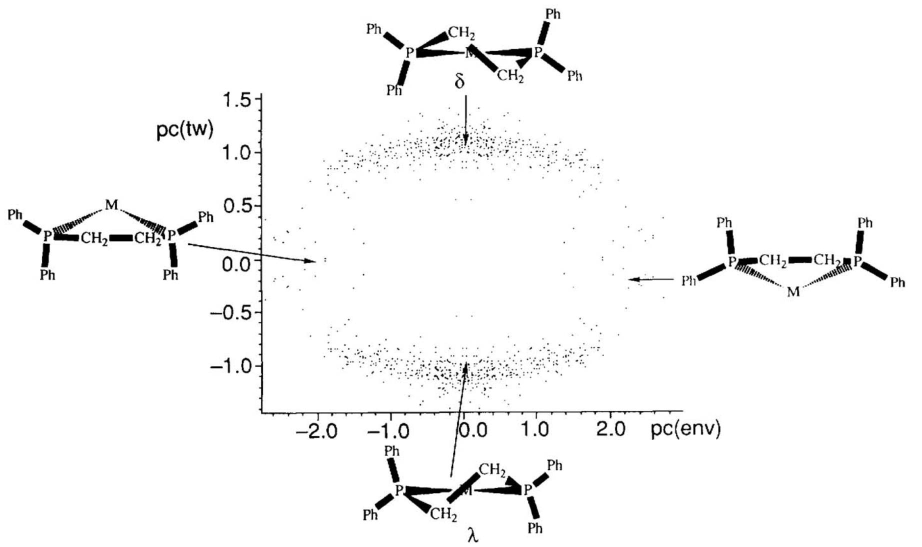


Figure 5 Scatterplot for two principal components obtained by pca of the M(dppe) five-membered ring torsion angle dataset. The principal components measure the degree of C_2 twist [pc(tw)] and C_s envelope [pc(env)] character in the conformation of any given M(dppe) fragment. The pseudo-rotation pathway interconverting δ and λ conformers *via* envelope intermediates is clearly visible.

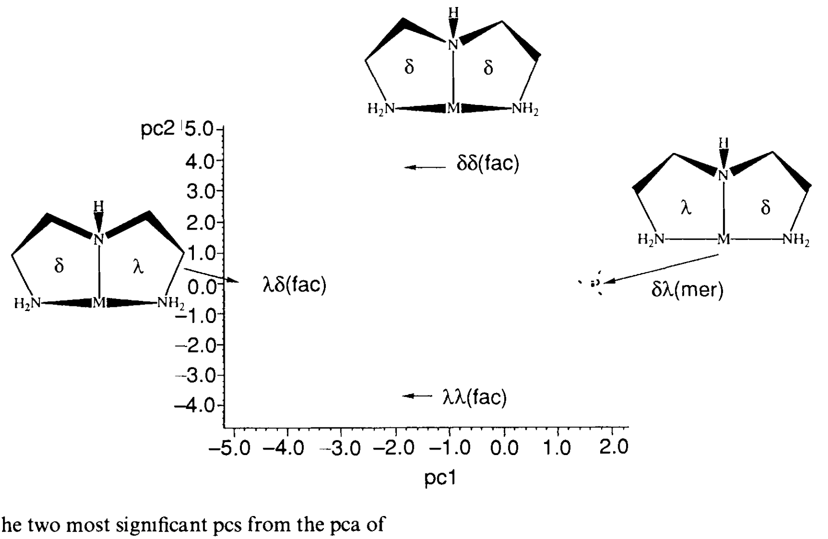


Figure 6 Pca scatterplot of the two most significant pcs from the pca of the ten intra-ring torsion angles of the dataset for M(dien). The four main conformer types and the corresponding coordination geometries are indicated.

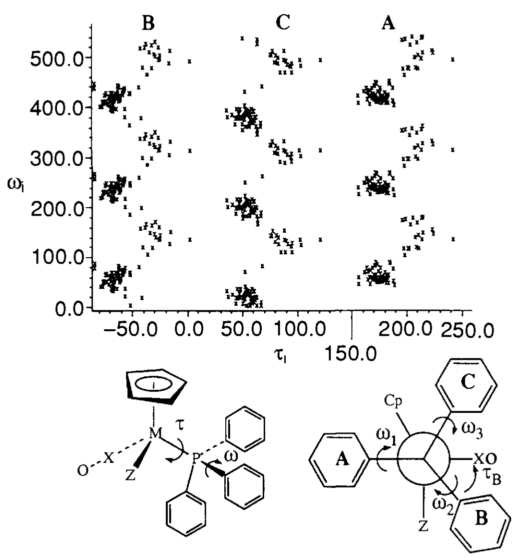


Figure 7 Scatterplot of torsion angles about the M−P (τ) and P−C (ω_1, ω_2, and ω_3) bonds for the chiral metal complexes [CpM(XO)Z(PPh$_3$)]. Note the vacant region of τ near 20, 140, and 260°, and the near continuous distribution of ω values.

image $\lambda\lambda$) or $\lambda\delta$ (which is *not* the mirror image of $\delta\lambda$) is observed for *fac*-M(dien) species.

In a different type of study,[9] not involving pca, we analysed the conformations of PPh$_3$ complexed to the asymmetric metal fragment CpM(XO)Z (X = N, C; Z = any ligand). Clear correlation of phenyl group and PPh$_3$ orientations (see Figure 7) and preferred conformations were observed, consistent with restricted rotation about the M−P bond but a relatively low barrier to P−Ph rotation. This proposition has been confirmed by NMR spectroscopic evidence, and is in good agreement with molecular mechanics treatments of this system.[18]

6 Reaction Pathway Analysis

Extension of PEH analysis has been taken to its logical conclusion in a series of outstanding studies by Bürgi and Dunitz and their co-workers. They applied the structure correlation method to a variety of inter- and intramolecular reactions.[19] The assumptions and hypotheses were essentially as described above except that in this context the reactions typically included bond formation and cleavage rather than conformational changes. In more recent extensions to this work, Bürgi has demonstrated the possibility of structure–energy correlations,[20] as well as structure–structure correlations (see discussion in reference 2 for example).

We have looked at two systems in which bond formation and cleavage information is available from study of collections of structural data. In each case we sought information to allow us to distinguish between alternative mechanisms. In the first we analysed possible pathways by which heteronuclear metal clusters containing Ru$_3$Au$_2$ fragments might rearrange.[21] In most known examples, these five atom fragments adopt a trigonal bipyramidal geometry with the two gold atoms in different sites, one axial and the other equatorial. From NMR studies it was known that the gold atoms in these clusters undergo rapid intramolecular site exchange with a modest energy barrier. The structural data available allowed a complete mapping of the trajectory followed in this exchange process and confirmed that a partial Berry pseudo-rotation path, in which a square pyramidal geometry is intermediate between two equivalent trigonal bipyramid cluster cores (see Figure 8), is followed. In contrast a number of other mechanisms could be ruled out as being inconsistent with the structural evidence. Formally the square pyramidal geometry has one less metal–metal bond (or short contact, since these species are surely not well described by localized two-electron metal–metal bonds) than does the trigonal bipyramidal species (8 compared with 9). In practice there is clear evidence of contraction of the metal–metal bonds around the open square face of the square pyramid apparently compensating for loss of bonding across the diagonal of this face.

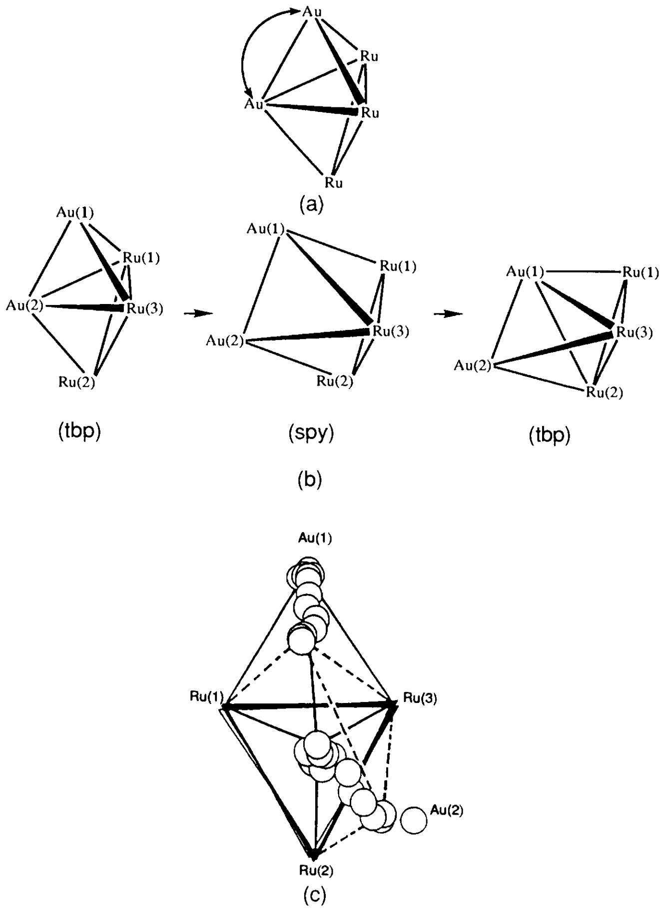


Figure 8 Metal site exchange in Ru$_3$Au$_2$ fragments. (a) Axial–equatorial exchange of gold atoms observed by NMR spectroscopy. (b) A partial Berry pseudo-rotation mechanism for this process. (c) The trajectories of the gold atoms during first half of the Berry-like process in (b), the Ru$_3$ triangles are superimposed and example tbp and spy (dashed) frameworks are shown.

In the second of our studies in this area we have sought to understand the means by which a carbon monoxide ligand moves between metal atoms in dinuclear and trinuclear arrays.[22] Such motion is of course exceedingly well known in the chemistry of metal clusters from NMR spectroscopy, and relevant structural evidence abounds. In a revealing study, Crabtree and Lavin[23] were able to identify a full range of structures to map out the CO exchange process in diiron species. In our work we too observed clear pathways for bridge–terminal exchange (see Figure 9), and have extended this to include CO ligands attached to triangular metal arrays. In this work one of our aims was to gain insight into CO migration across the cluster and by implication into related process on metal surfaces, in which of course there are many triangular M$_3$ units (*e.g.* as on close-packed metal surfaces). One question that might be answerable in this way is how does CO move from a terminal site into a triply bridging position? The two extreme potential routes would be (i) directly from the terminal site through intermediate geometries of C_s symmetry [see (a), Figure 10], and (ii) through μ_2-intermediates [see (b) and (c), Figure 10]. Inspection of the distribution of CO geometries above the metal triangle (see Figure 10), with due allowance for the symmetry of the parameter space, suggested that the path through (c) and (b) is the more likely, although there is a small number of structures in the region close to the intermediate geometries appropriate to path

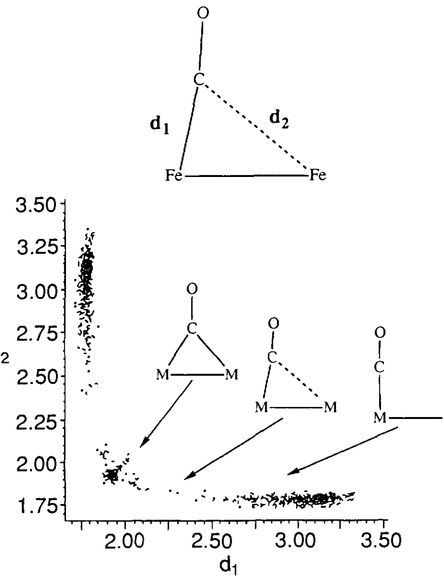


Figure 9 Distribution of Fe$-$C distance d_1 and d_2 in $Fe_2(CO)$ fragments in the CSD. The narrow band of values observed corresponds to the reaction pathway for terminal to μ_2-CO site exchange.

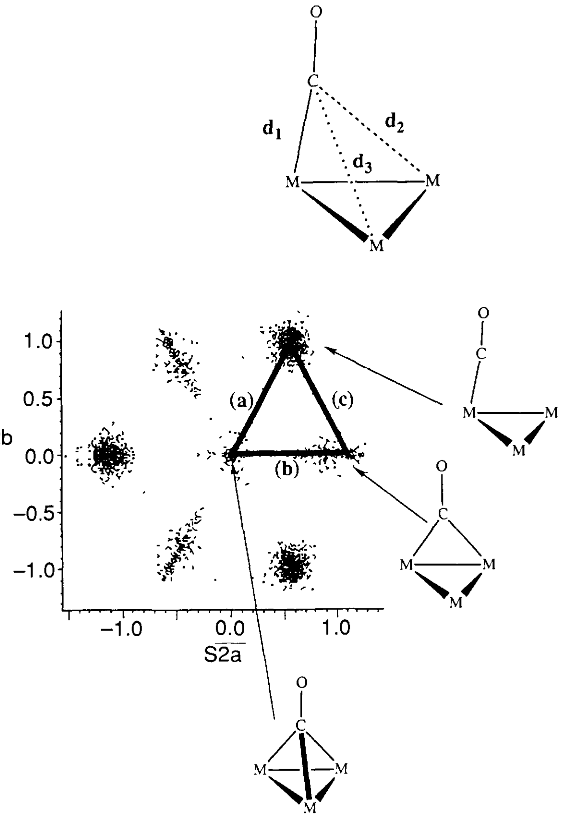


Figure 10 Scatterplot (*S*2b *vs.* *S*2a) for $M_3(CO)$ fragment geometries ($M_3 = Fe_3$, Ru_3, Os_3, Co_3, Rh_3, or Ir_3). $S2a = (2d_1 - d_2 - d_3)/\sqrt{6}$; $S2b = (d_2 - d_3)/\sqrt{2}$. Pathways for direct terminal to μ_3-CO site exchange (a), μ_2- to μ_3-CO site exchange (b), and terminal to μ_2-CO site exchange (c), are indicated.

(a). Another aspect investigated in this study was the effect of carbonyl ligand position on the dimensions of the polymetal unit. For example, in $Ru_2(CO)$ species the range of Ru$-$Ru (single) bond distances was greatly reduced in those cases where the carbonyl is close to symmetrically bridging. However, when the CO was terminal Ru$\cdots$Ru distances spanned the range from 2.6 to 3.2Å; when the carbonyl was bridging the range was reduced to 2.6—2.85Å, and as a result the mean Ru$-$Ru distance for this class of structures was substantially lower than for the entire set of structures. Similar effects are observed for the M_3 system. The limits on the range of metal–metal distances presumably arise because of limit on the range of tenable M$-$C$-$M angles in bridging carbonyls.

The M$-$C distances for μ_2-CO ligands were only slightly increased over the terminal values (by *ca.* 12%, see Figure 9). This is considerably smaller than for other bridge-terminal exchange processes studied in this way, *e.g.* O$-$H$\cdots$O to O$\cdots$H$-$O,[19] for which the symmetrical intermediate has O$-$H bonds *ca.* 25% longer than in the terminal cases. These latter variations in interatomic distances were quantitatively explained as arising from the conservation of bond order around the O$-$H$\cdots$O system. The small M$-$C bond length extension is therefore consistent with this same principle, but only if some loss of bond order occurs elsewhere in the system (*i.e.* in the C$-$O bond). Of course this is precisely what is expected given the lower values of ν(CO) and associated increases in C$-$O bond lengths for bridging as compared with terminal carbonyl ligands.

7 Conclusions

In many cases the determination of a crystal structure is an end in itself, and a crucial element in a research project. This review has indicated some ways in which a broader view of the structure, and in particular, its variability under perturbation, may be obtained. In part, this type of study seeks to offer some insight into the eternal conundrum of crystallography: 'so that is the structure in the crystal – but how different/similar is it in solution?' While crystallographic results alone cannot answer this question, we can learn something by seeing a fragment in a wide variety of crystal environments. The different fields perturb the fragment geometry and may be viewed as mimicking the variety of local fields that the same fragment would experience over time in a solution. Clearly this is a much simplified view and is not quantitative. However, we can safely say that a fragment that shows very little variation in geometry across a wide range of crystal (and/or molecular) environments will spend essentially all its time in such a geometry while in solution.

Collections of structures may be of use in other ways too. There are clearly many physical observables against which one might choose to calibrate a molecular modelling method, including structural, thermodynamic, and spectroscopic information. Of these, perhaps the most reliable and abundant source of data is from three-dimensional molecular structures revealed by crystallography. Clearly one can model the molecular geometry found in an individual crystal structure and thereby assess the efficacy of a modelling procedure. However in most calculations the procedure predicts the properties of an isolated (gas-phase) molecule, usually at rest. In contrast the crystallography data refer to a molecule in a crystal field and undergoing vibrational motion. Two strategies might be considered to make a more appropriate comparison between the model and crystallographic results: (1) to allow explicitly for the symmetry (including translational symmetry) of the solid state structure in the modelling procedure; (2) to place emphasis on the ability to reproduce the range of geometries adopted by a molecular or submolecular fragment in a large number of crystal structures. The aim of this latter approach is to find not just the calculated minimum energy molecular or fragment configuration, but also to examine its softness towards perturbation from whatever source (and by implication to learn something of the fragment's potential energy hypersurface). Clearly such studies are of particular utility when molecular flexibility or reactivity are of importance. This latter approach has been successfully used by Klebe[24] and others. More direct relationships between molecular structures and their energetics have been probed by Bürgi and his co-workers in structure–energy correlation studies.[2,20]

Developments in the field of structural systematics may in a sense be largely limited by the imagination of the practitioners since there is such an enormous quantity of crystallographic data to hand. In the types of studies described above, the structure–structure correlation methodology can be extended to gain more insight into how structural parameters interact. For example, we might study how the conformation of ring systems affects, or is affected by, other aspects of geometry (*e.g.* large M—M distances are associated with rigorously envelope ring conformation in the $M_2(\mu\text{-dppm})$ system).[15] This approach is clearly linked to the desire to be able to engineer molecular structure so as to gain control of reactivity, catalytic, and other properties of the species. What is missing in much of this type of work is an understanding of the relationship beween the molecular geometry and the reactivity or other property of interest. Such studies will require a substantial body of physico-chemical data for correlation with the structural data, but there is clearly an opportunity to aim for quantitative structure–activity relationships of the sort that have been so successfully exploited in pharmaceutical chemistry.

In the studies described above it has become clear that it is necessary to use techniques (frequently statistical) which are unfamiliar to many chemists in order to analyse and make the best use of the vast quantities of data in the CSD. These include the application of principal component analysis and the exploitation of the symmetry of parameter space, as described above. Substantial progress has been made in introducing statistical techniques[13,14,25] and adapting them to deal with the symmetry and periodicity properties of the parameter spaces of interest to structural chemists. Nevertheless it is clear that there remains much to be learnt about the most effective ways of extracting chemical understanding from structural data, and that it is through statistical methods that most of this progress is likely to be made.

In addition to technical developments, the prospects for work in this area are perhaps brightest in the field of extramolecular chemistry – the interactions between molecular species in the crystalline state. Crystallography is, of course, by definition a unique source of information in this respect. The scope of the potential contribution of this approach to the study of solid-state reactivity, surface chemistry, and the field of molecular recognition in molecular inorganic chemistry is tremendous and has only begun to be recognized and exploited.[26]

The emphasis in this article has been on the use of empirical correlations between structural parameters. In fact, of course, any useful chemical interpretation of this information is obtained by adding a considerable patina of theory and hypothesis to the empirical observations. The reader should be aware of this. One particular consequence is that there may be several hypotheses that serve to rationalize a given set of observed correlations. This is no weakness of the approaches outlined, but merely another reminder of the way science is.[5]

To sum up, our aim is to understand first what shapes molecules have, then to think about why they have those shapes, and finally to explore the chemical consequences of their shape. The abundance of molecular structure data available, and current developments of methodology and technology, afford extraordinary opportunities for new structural science.

Acknowledgements. I thank the Ciba-Geigy Foundation for the award of a Senior Research Fellowship, Professor D. Braga and his colleagues at the Dipartimento di Chimica 'G. Ciamician', Università di Bologna for their hospitality during a sabbatical stay in Bologna during which time this article was drafted, my colleagues and co-workers at the University of Bristol, and the many chemists and crystallographers who synthesized and determined the structures of the compounds discussed.

8 References

1 F. H. Allen, O. Kennard, and R. Taylor, *Acc. Chem. Res.*, 1983, **16**, 146; F. H. Allen, J. E. Davies, J. J. Galloy, O. Johnson, O. Kennard, C. F. Macrae, E. M. Mitchell, G. F. Mitchell, J. M. Smith, and D. G. Watson, *J. Chem. Inf. Comput. Sci.*, 1987, **31**, 187.

2 See H.-B. Bürgi, in 'Perspectives in Coordination Chemistry', ed. A. F. Williams, C. Floriani, and A. E. Merbach, Verlag Helvetica Chimica Acta, Basel, 1992, p.1; T.P.E. Auf der Heyde, in 'Structure Correlation', ed. H.-B. Bürgi and J. D. Dunitz, VCH, Weinheim, in the press.

3 F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, and R. Taylor, *J. Chem. Soc., Perkin Trans 2*, 1987, S1.

4 A. G. Orpen, L. Brammer, F. H. Allen, O. Kennard, D. G. Watson, and R. Taylor, *J. Chem. Soc., Dalton Trans.*, 1989, S1.

5 K. R. Popper, 'Conjectures and Refutations', Routledge and Kegan Paul, London, 1963.

6 A. G. Orpen and N. G. Connelly, *J. Chem. Soc., Chem. Commun.*, 1985, 1310; A. G. Orpen and N. G. Connelly, *Organometallics*, 1990, **9**, 1206.

7 See, *e.g.*, R. P. Aggarwal, N. G. Connelly, M. C. Crespo, B. J. Dunne, P. M. Hopkins, and A. G. Orpen, *J. Chem. Soc., Dalton Trans.*, 1992, 655.

8 B. J. Dunne, R. B. Morris, and A. G. Orpen, *J. Chem. Soc., Dalton Trans.*, 1991, 653.

9 S. E. Garner and A. G. Orpen, *Acta Crystallogr., Sect. A*, 1990, **A46**, C360; *J. Chem. Soc., Dalton Trans.*, 1993, 533.

10 H. B. Bürgi, A. Raselli, D. Braga, and F. Grepioni, *Acta Crystallogr., Sect. B*, 1992, **B48**, 428.

11 J. D. Dunitz and H.-B. Bürgi, *Acta Crystallogr., Sect. B*, 1988, **B44**, 445.

12 R. W. Alder, C. M. Maunder, and A. G. Orpen, *Tetrahedron Lett.*, 1990, **31**, 6717.

13 H. B. Bürgi and K. Chandrasekhar, *J. Am. Chem. Soc.*, 1983, **105**, 7081; L. Nørskov-Lauritsen and H. B. Bürgi, *J. Comp. Chem.*, 1984, **6**, 216.

14 (*a*) F. H. Allen, M. J. Doyle, and R. Taylor, *Acta Crystallogr., Sect. B*, 191, **B47**, 29, 41, and 50; (*b*) F. H. Allen, M. J. Doyle, and T. P. E. Auf der Heyde, *Acta Crystallogr., Sect. B*, 1991, **B47**, 412.

15 D. A. V. Morton and A. G. Orpen, *J. Chem. Soc., Dalton Trans.*, 1992, 641.

16 D. Cremer and J. A. Pople, *J. Am. Chem. Soc.*, 1975, **97**, 1358.

17 S. E. Garner and A. G. Orpen, unpublished results.

18 S. G. Davies, A. E. Derome, and J. P. McNally, *J. Am. Chem. Soc.*, 1991, **113**, 2854.

19 H. B. Bürgi and J. D. Dunitz, *Acc. Chem. Res.*, 1983, **16**, 153; H. B. Bürgi, *Inorg. Chem.*, 1973, **12**, 2321.

20 H. B. Bürgi and K. C. Dubler-Steudle, *J. Am. Chem. Soc.*, 1988, **110**, 4953, 7291.

21 A. G. Orpen, and I. D. Salter, *Organometallics*, 1991, **10**, 111.

22 L. Brammer, N. S. Dhillon, D. A. V. Morton, K. A. MacPherson, and A. G. Orpen, to be submitted to *Inorganica Chimica Acta*.

23 R. H. Crabtree and M. Lavin, *Inorg. Chem.*, 1986, **25**, 805.

24 G. Klebe, *Struct. Chem.*, 1990, **1**, 597.

25 R. Taylor, *J. Mol. Graphics*, 1986, **4**, 123.

26 See for example, D. Braga and F. Grepioni, *Organometallics*, 1992, **11**, 711, 1256.